You can go your own way: Effectiveness of participant-driven versus experimenter-driven processing strategies in memory training and transfer

Kristin E. Flegal[1] and Cindy Lustig[2]

[1] Institute of Neuroscience and Psychology, University of Glasgow, Scotland, UK

[2] Department of Psychology, University of Michigan, Ann Arbor, MI, USA

Keywords:

AGING, MEMORY, TRAINING, TRANSFER, COGNITIVE REHABILITATION

Abstract

Cognitive training programs that instruct specific strategies frequently show limited transfer. Open-ended approaches can achieve greater transfer, but may fail to benefit many older adults due to age deficits in self-initiated processing. We examined whether a compromise that encourages effort at encoding without an experimenter-prescribed strategy might yield better results. Older adults completed memory training under conditions that either 1) mandated a specific strategy to increase deep, associative encoding, 2) attempted to suppress such encoding by mandating rote rehearsal, or 3) encouraged time and effort towards encoding but allowed for strategy choice. The experimenter-enforced associative encoding strategy succeeded in creating integrated representations of studied items, but training-task progress was related to pre-existing ability. Independent of condition assignment, self-reported deep encoding was associated with positive training and transfer effects, suggesting that the most beneficial outcomes occur when environmental support guiding effort is provided but participants generate their own strategies.

Introduction

Older adults commonly report that they don’t remember as well as when they were younger, and their firsthand experience is consistent with data from both cross-sectional and longitudinal studies of age-related cognitive change (Lustig & Lin, in press; Rönnlund et al., 2005; Spencer & Raz, 1995). One factor contributing to memory decline in normal aging is that older adults often fail to spontaneously engage deep, associative encoding processes that facilitate later memory (Craik & Byrd, 1982; Craik & Rose, 2012). However, research has consistently shown that providing instructions and environmental support to encourage such processing can change older adults’ memory performance to more closely resemble that of young adults (e.g., Braver, Gray, & Burgess, 2007; Craik & Byrd, 1982; Froger et al., 2012; Logan et al., 2002; Naveh-Benjamin, Brav, & Levy, 2007; Paxton et al., 2006). These findings are important in indicating that effortful cognitive abilities are not entirely lost with advancing age, but are frequently latent in the absence of external support. Furthermore, they hint at individual differences in strategy use under unsupervised learning conditions, illustrating the wide variety of predispositions and preferences that older adults bring to bear on memory tasks.

Older adults’ failure to self-initiate effective memory strategies may reflect a *production deficit*, in which strategies are accessible but not spontaneously produced (Dunlosky & Hertzog, 1998; Kausler, 1994; Verhaeghen & Marcoen, 1994), or a *utilization deficit*, in which strategies are produced but not successfully employed (Dunlosky & Hertzog, 1998; Dunlosky, Hertzog, & Powell-Moman, 2005; Jones et al., 2006) – both potentially in addition to, or interacting with, age-related declines in processing capacity (Jones, et al., 2006; Salthouse, 1996). Both strategy-based explanations are consistent with findings that older adults can and do benefit from experimenter-prescribed instructions in the context of memory training (Rebok, Carlson, & Langbaurn, 2007; Verhaeghen, et al., 1992). Additionally, there is evidence to suggest that training with effective experimenter-provided strategies may help older adults to overcome difficulties with self-initiation when external support is withdrawn: Kirchhoff and colleagues (2012a) found that two sessions of training with semantic encoding strategies subsequently increased older adults' self-initiated use of such strategies during unsupported, intentional encoding – as well as their post-training recognition memory performance.

Without explicit guidance or external support, however, older adults are often less likely than young adults to self-report using effective memory strategies spontaneously (Dunlosky & Hertzog, 2001; Touron & Hertzog, 2004). And even with instruction, older adults may be less likely than young adults to apply the experimenter-mandated strategy correctly (Verhaeghen & Marcoen, 1996). Thus, specific mnemonics provided by an experimenter can help some older adults overcome a production deficit, but at the risk of creating a utilization deficit for those with lower initial ability. Although experimenter-driven strategy instruction may initially help to compensate for ability differences, some individuals' limited capacity to benefit from training with a particular strategy can subsequently magnify differences in memory performance (Lövdén et al., 2012). An understudied question in cognitive training research is whether environmental support itself (such as conditions that encourage sufficient time and effort at encoding) can prompt older adults to engage individualized strategies that will benefit memory performance. This is an important consideration for the development of training interventions to maintain or enhance cognitive fitness late into the lifespan, as some evidence suggests that older adults' self-generated processing strategies can be at least as effective as experimenter-provided mnemonics (Baltes, Sowarka, & Kliegl, 1989; Derwinger et al., 2003; Hill, Allen, & Gregory, 1990) and may produce long-lasting benefits which are less dependent on environmental support (Derwinger, Stigsdotter Neely, & Bäckman, 2005).

The present study tests the hypothesis that encouraging older adults to spend sufficient time and effort encoding information for better memorability will improve memory performance, especially for individuals who might otherwise fail to self-initiate effective encoding strategies. Facilitating the use of deep encoding is predicted to both increase gains on the training task and transfer to other tasks, but only those with semantically-based and/or integrative components; these include measures of real-world memory. While many training interventions emphasize bolstering the cognitive processes that weaken with age, the possibility of capitalizing on the processes that remain intact – but inefficiently used – has often been overlooked (Park et al., 2007). Explicit strategies or mnemonics provided by an experimenter may aid participants who otherwise would not have thought to employ them, but this one-size-fits-all approach is insensitive to individual differences in existing cognitive strengths and weaknesses, which is a particular concern for low-ability individuals (Calero & Navarro, 2007; Hill et al., 1989; Verhaeghen & Marcoen, 1996). For example, training with verbal encoding strategies may be suboptimal for individuals who do not already possess strong verbal skills (Yesavage et al., 1988). This problem is compounded by the finding that while strategy-based training can improve older adults' performance on a target task (e.g., Ball et al., 2002; Rebok & Balcerak, 1989), the benefit frequently fails to transfer to untrained tasks (see Lustig et al., 2009, for a review and further discussion). However, sufficient external support in the absence of specific encoding instructions might foster the self-generation of effective memory strategies tailored to participants' individual preferences and pre-existing strengths, and this approach may be more likely to promote transfer to real-world situations.

One way to provide environmental support for encoding is by allowing extended time to produce or utilize effective strategies. For example, Thompson and Kliegl (1991) found that recall accuracy could be equated between age groups when encoding times were on the order of three times longer for older adults than young adults. However, research has also demonstrated that the length of time available in intentional encoding tasks interacts with initial ability in predicting older adults' self-generated strategy use (Craik & Rabinowitz, 1985). Similarly, unlimited encoding time may not reduce age differences in memory performance, because under self-paced conditions, older adults tend to allocate less time in a study phase than young adults (Dunlosky & Connor, 1997; Murphy et al., 1981) – perhaps due to differential use of information from metacognitive monitoring (Dunlosky, Kubat-Silman, & Hertzog, 2003). Comparable effects are revealed by individual differences within, not only between, age groups: In a memory-training study with unconstrained encoding time, Bissig and Lustig (2007) found that older adults who spent proportionally less time on study phase trials than test phase trials showed poorer performance on the training task. Thus, simply granting older adults as much study time as they choose is not enough to overcome their deficits in self-initiating effortful memory processes, indicating that additional environmental support may produce the most beneficial outcomes. In fact, Froger et al. (2012) found that providing a high level of environmental support in an associative memory task, by giving instructions for encoding strategies along with information about their effectiveness, led older adults to allocate more study time on more difficult trials under self-paced conditions.

On the basis of previous literature identifying encoding processes as an attractive target for training in older adults, and an initial study revealing the extent to which differences at encoding accounted for training gains under open-ended conditions (Bissig & Lustig, 2007), we designed a memory training intervention which imposed generous study times for all participants and manipulated encoding instructions. Based on the repetition-lag procedure developed by Jennings and Jacoby (2003), the present study examined training and transfer effects in conditions that either 1) mandated a deep, associative encoding strategy, 2) attempted to suppress such encoding by mandating rote rehearsal, or 3) encouraged effort towards encoding (by enforcing study times) but allowed participants to choose their own strategies. Our initial hypotheses (see Lustig & Flegal, 2008) were formed around the consequences of enforcing strategies believed to either benefit memory performance (integrative encoding) or to suppress effective encoding processes (rote rehearsal). We anticipated that instructing one group of older adults to use strategies reported by the most successful participants in our earlier training study might help them to overcome difficulties with self-initiation and potentially minimize the influence of pre-existing ability. The condition in which no specific strategy was provided controlled for the amount of encoding time but was otherwise analogous to the unconstrained setting of our earlier study (Bissig & Lustig, 2007), and we initially expected that a number of these participants would likewise struggle to self-initiate effective encoding strategies without explicit guidance.

Although not a focus of our original predictions, individual differences emerged as an important factor in our analyses, and we discovered that environmental support in the form of fixed encoding time benefitted participants in all three experimental conditions (and was sufficient in itself to diminish age and ability differences in training task performance). We also found that pre-existing ability influenced training gains even when encoding instructions were experimenter-controlled, and that a variable at least as important as how “effective” a memory strategy was deemed to be is whether a memory strategy was mandated or self-selected. There were hints of these effects in our preliminary data (Lustig & Flegal, 2008), and they are consistent with previous research associating self-generated strategies with superior memory performance in older adults (Derwinger, et al., 2005; Derwinger, et al., 2003; Hill, et al., 1990). Of interest, earlier studies have reported that older adults benefit as much from extended practice on a task involving attention to contextual cues as from explicit strategy training (Paxton et al., 2006), and that older adults in an enforced encoding time condition perform better on a serial recall task than a self-paced group instructed to maximize accuracy by taking as much encoding time as necessary, and even outperform a self-paced group given explicit strategy training to improve their accuracy (Murphy et al., 1981). Such outcomes are concordant with findings from the present study which suggest that enforcing ample time-on-task may provide a greater benefit for increasing self-initiated processing than using that same amount of time to train with an experimenter-prescribed strategy.

Method

Participants

Ninety healthy older adults (n = 30 per group; 65-92 years of age; 61 female) were assigned to the three training conditions, with the groups matched closely in age, education, and gender (see Table 1 for demographics). Within each group, participants were stratified by age in 5-year bins (i.e., an approximately equal number of participants age 65-69, 70-74, 75-79, etc.). All participants were screened for medical or psychological conditions that could influence performance, and had Mini Mental State Evaluation scores (MMSE; Folstein, Folstein, & McHugh, 1975) scores above 24 (mean = 28.4). The study was approved by the University of Michigan Institutional Review Board, and written informed consent was obtained from all participants. None of the participants withdrew from the study prior to completing all eight sessions.

Materials and procedure

Scheduling. Each participant completed eight study visits scheduled over the course of three weeks. Day 1 included informed consent procedures, a health and demographics questionnaire, dementia screening measures (MMSE and Short Blessed Test; Katzman et al., 1983), the Extended Range Vocabulary Test (ERVT; Educational Testing Services, 1976) as a measure of verbal ability, baseline (pretest) administration of potential transfer tasks, and brief practice with the training task to familiarize participants with the encoding instructions and time constraints. The last day (Day 8) included posttest administration of the transfer tasks and a questionnaire about strategies used in the training task. This questionnaire was administered at the very end of the last study visit.

Training task. The training task was a modified version of the repetition-lag procedure developed by Jennings and Jacoby (2003; see also Jennings et al., 2005). The base procedure consists of 28 study-test “cycles”. In each cycle, participants first study 30 words, presented one at a time. The test phase consists of a yes/no recognition test that includes the 30 studied words as well as 30 unstudied words (lures). The unstudied words repeat within the test list (classified as “new” on their first presentation and “repeated” on their second presentation), requiring participants to discriminate items that are familiar because they were encountered in the study phase from those that are familiar because they were previously-presented lures from earlier in the test phase. A feedback screen appeared after each response, indicating accuracy (correct or incorrect) and trial type (studied, new, or repeated).

The level of difficulty for the retrieval test was dynamically adapted to individual performance by increasing the number of items (“lag”) between lure repetitions in the test phase once criterion performance was reached at the current level. At each level, half of the lure repetitions occurred at a short lag (few items in between repetitions) at which the participant could perform well. These trials were included to help to maintain confidence and motivation. The other half occurred at a long lag (more items in between repetitions), and were used to challenge performance. The possible lag-interval combinations were 1&2, 1&3, 2&4, 2&8, 4&12, 4&16, 8&20, 8&24, 12&28, 12&32, 16&36, and 16&40. All participants started the training task at the easiest level, at which half of the lures repeated after only one intervening word and the other half repeated after two intervening words (i.e., lag level 1&2; see Figure 1A). The criterion for advancing to the next lag level was set at 96% correct rejections of long-lag repeated lures for levels up to 2 and 8 (i.e., the 4th level) and relaxed to 93% for higher levels. Once a participant reached the maximum level (16&40), she or he continued working at that level for the remaining sessions. Participants completed four study-test cycles on each of the seven days of training, for a total of 28 training cycles.

Study and test words were chosen from the English Lexicon Project (Balota et al., 2007) and had a mean length of 5.76 letters and mean frequency of 20,487 out of 131 million. Length and frequency were balanced across lists and across conditions (studied, unstudied-short-lag, unstudied-long-lag). Each word was presented in large (32-point Arial) black-on-white font in the center of a computer screen. E-prime software (Psychology Software Tools, Pittsburgh, PA) was used for stimulus presentation and response collection (via key press).

The major differences between the training procedure used here (see also Lustig & Flegal, 2008) and the procedure originally designed by Jennings and Jacoby (2003) occurred at the study phase. These differences included the encoding times used in the study phase of the procedure (2 seconds per word in the original study; 14 seconds per word here), and the assignment of participants to one of three encoding conditions (Integrated Sentences, Strategy Choice, Enforced Rehearsal). The longer encoding time used here was chosen on the basis of the encoding times used by good performers in a previous experiment (Bissig & Lustig, 2007) and on the time needed by a separate, pilot group of older adult participants to implement the Integrated Sentences condition.

In the Integrated Sentences condition, participants were instructed to make a sentence out of each word and (for all but the first word in the list) the word that had just preceded it. Participants were to say the sentences aloud, and they were recorded to ensure compliance and for later content analysis. In the Strategy Choice condition, participants were instructed to think about the meaning of each word presented during encoding in any way they would like, but no explicit strategy was specified. In the Enforced Rehearsal condition, participants were instructed to repeat each word out loud at a fixed pace of once every 2 seconds, guided by a counter on the computer screen. A preliminary report of data from a subset (n = 16 per group) of participants in the Integrated Sentences and Strategy Choice conditions appeared in Lustig and Flegal (2008). The present results generally corroborate that early report and add new findings; any discrepancies are highlighted in the Results section below.

Transfer tasks. A battery of transfer tasks (see Table 2) was administered before the first day of training and immediately following the last session. These tasks were used to test the transfer of training benefits to untrained tasks, to help identify which processes were being trained by the intervention, and to assess whether the trained processes and the effectiveness of training differed across encoding conditions. Across all transfer tasks, items were carefully screened to avoid overlap with the training lists and with the stimuli used in other transfer tasks. Alternate forms were used at pre- versus post-test for all measures except for the Trail-Making Test, for which they do not exist (see below).

Some of the tasks were hypothesized to show transfer effects because they emphasized semantically-based and/or integrative processes that were expected to improve if the training task primarily increased deep, associative processing at encoding. One such measure was a shopping-list memory task, in which participants studied a list of 15 individually presented words representing typical grocery items (e.g., “eggs”, “lettuce”) and were then given a self-paced old/new recognition test for the 15 studied items and 30 unstudied lure items. On Day 8, half of the “unstudied” items were words that had previously been studied items on the shopping list on Day 1, and half were completely new. On posttest administration, participants were instructed to identify as “old” only those words that had been on the list that day, and to call all other words “new” even if some might have been seen on an earlier day. This manipulation was designed to assess participants’ ability to resist proactive interference from pretest to posttest, when previously studied items would be (potentially) familiar but no longer relevant. Another measure on which we predicted positive transfer was a face-name association task, in which participants studied 10 face-name pairs (using digitized black and white photographs, with self-paced encoding and instructions to make a sentence on each trial that connected the name with the face). Participants were then given a two-stage memory test, in which they were first presented with each face and asked to recall the name that had been paired with it; if they could not recall the name, they were given the correct name and a lure name and asked to identify the correct one.

Other tasks were used as “negative controls” to test the hypothesis that general practice or engagement and stimulation as a result of participation in the training program might lead to performance improvements without regard to which processes were targeted for training. These tasks did not emphasize the semantically-based or integrative processes targeted by our training procedure, and so were predicted to show little or no improvements. These measures included the Pattern Comparison Test (Salthouse & Babcock, 1991), a common measure of cognitive speed; the Trail-Making Test (Armitage, 1946), which measures cognitive speed (Part A) and executive function (Part B); and pattern and word versions of a self-ordered pointing test (SOPT; Attneave & Arnoult, 1956), a measure of working memory. Our versions of the SOPT consisted of 16 words or 16 patterns arranged in a 4 × 4 grid. There were 16 pages for each test, and the 16 items for that test were arranged differently on each page. On each page, the participant's task was to point to an item to which he or she had not previously pointed.

Another purpose of the transfer task battery was to assess the potential impact of our training procedure on real-world memory. In addition to measures such as the shopping-list memory task and face-name association task, which were designed to simulate the type of memory tasks that older adults encounter in daily life outside the laboratory, participants completed the 35-item Everyday Memory Questionnaire (EMQ; Sunderland, Harris, & Baddeley, 1983) on the first visit and prior to each of their following visits. Participants were asked to indicate how many times within the last 24 hours they had committed each of the memory errors listed on the questionnaire. Participants also completed the Memory Self-Efficacy Questionnaire (MSEQ-4; Berry, West, & Dennehey, 1989) at pretest and posttest. Participants were asked to rate their confidence in performing different memory tasks (e.g., remembering parts of a story or items on a shopping list) at different levels of difficulty (two items, eight items, and so forth). Any inconsistencies observed between EMQ and MSEQ responses could help to rule out placebo effects or factors not directly related to the intervention as explanations for self-reported improvements in everyday memory, because fewer EMQ errors should be correlated with greater confidence in memory ability if the effect were driven by expectancy of post-training improvement (a summary of the baseline EMQ and MSEQ scores from this sample is reported in Ossher, Flegal, & Lustig, 2013).

Two additional transfer tasks were administered only at the end of training (i.e., no pretest) because of the likelihood that knowledge of the test format would influence the strategies participants engaged at encoding. Immediately following the final study-test cycle on the last day of training, participants were given a surprise recognition test for unstudied lures from the final training test list (see Figure 1B). This lure test included 15 studied words that appeared on the preceding test, 15 words that had served as lures on the preceding test (i.e., “previous lures”), and 30 completely new words. Participants were now instructed to identify as “old” any word that had appeared in the final test session, regardless of whether it was originally a studied word or if it had been a “new” word (a previous lure item) to be rejected as unstudied at the time. Finally, in a word- and source-memory task, 30 words were auditorily presented, half in a male voice and half in a female voice (counterbalanced across subjects), pseudorandomly intermixed. The time to advance to the next study trial was self-paced, but the auditory stimuli themselves were relatively brief in duration (as they were single words spoken at a normal speed) and could not be repeated. This study phase was followed by a visually presented 60-item recognition test in which participants indicated whether words were old or new and, if old, whether they had been spoken in a male or female voice.

Results

We conducted between-subjects ANOVAs to evaluate the effects of experimenter-prescribed condition and self-reported encoding strategy (see “Encoding condition adherence”, below). For the training task, we analyzed 1) the maximum lag level achieved by the last day of training, to index outcomes from the study phase manipulation, 2) response times (RTs) in the test phase of the training task, to examine how retrieval processes were affected by training condition and item type (studied item, new lure, repeated lure), and 3) the amount of time that participants spent processing feedback screens. For the transfer tasks, we used mixed design ANOVAs with encoding condition or strategy as the between-subjects factor and assessment occasion (pretraining, posttraining) as a repeated measure. For both the training task and transfer task, results are first presented for the data analyzed by experimenter-assigned encoding condition (Integrated Sentences, Strategy Choice, Enforced Rehearsal), and then by participant-chosen encoding strategy (shallow, deep). Degrees of freedom reported below vary for some of the transfer tasks which were not completed by all 90 participants. The Greenhouse-Geisser sphericity correction was applied as needed to the statistical values reported below, with degrees of freedom rounded to integer values for ease of reading.

<u>Demographics and training-task performance</u>

Participants in the three encoding conditions did not differ in verbal ability (ERVT), speed (Pattern Comparison Test), years of education, or dementia scales (Table 1; all *p*s > .17 for main effects of Condition [Integrated Sentences, Strategy Choice, Enforced Rehearsal]). They also did not differ in the maximum lag level achieved by the end of the training sessions, nor in the number of sessions required to achieve the maximum possible lag, both Fs < 1.

As previously described (Bissig & Lustig, 2007; Lustig & Flegal, 2008), we ranked training-task performance (1 = *best*, 30 = *worst*) within each condition according to the highest lag level at which participants reached criterion performance, with ties between participants who reached the same maximum lag level resolved by granting the better (lower) rank to whichever participant reached criterion in the earlier training session. Remaining ties were broken by assigning the better rank to whichever participant had the higher overall accuracy for correct rejection of repeated lures. For example, a participant who reached criterion performance for the maximum lag level in Session 15 would receive a better rank than two other participants who both reached it in Session 21, and between those two participants, a better rank would be assigned to the one for whom overall repeated lure accuracy was 90% than the one for whom it was 84%.

In a previous study in which encoding was self-paced and unconstrained (Bissig & Lustig, 2007; “Open-Ended” condition in Figure 2), we found that age was associated with training rank, whereas verbal ability was not. This suggested that advanced age was associated with a reduced likelihood of self-initiating effective encoding, whereas letting participants choose which method(s) they might use allowed self-initiators with lower verbal skills to use an approach that worked better for them. The contrast between those results and the present study further support this hypothesis: Enforcing encoding time diminished the association between poorer training-task performance (indexed by the ranking variable) and older age for all three of the encoding conditions used here (all *r*s < .24, all *p*s > .22)[1]. Performance in the Integrated Sentences condition was significantly correlated with vocabulary and years of education, whereas this was not the case for the Strategy Choice condition; the Enforced Rehearsal condition showed a weaker and nonsignificant relation in this direction (Figure 2, Rows 1 and 2). This suggests that participants who had strong verbal skills to begin with were best able to benefit from the compulsory verbal strategy used in the Integrated Sentences condition. In contrast (when paired with the environmental support of long encoding times), the flexibility of the Strategy Choice instructions and the potential for the Enforced Rehearsal instructions to be augmented with other, covert strategies (see below) appears to have allowed participants in those conditions to engage encoding processes tailored to their personal strengths and preferences, thus diminishing the predictive power of pre-existing ability.

<u>Encoding condition adherence</u>. We were initially surprised to find that the Enforced Rehearsal group performed as well on the training task as did the other two groups. This condition was originally designed to serve as a control condition that would suppress deep encoding processes and mimic the strategies reported by the least successful participants in our previous study (Bissig & Lustig, 2007). Examination of the post-training strategy questionnaire revealed that many participants in the Enforced Rehearsal condition were covertly engaging deep encoding strategies while also completing the overt rehearsal task (e.g., “I made associations to the words wherever possible”; “I related each word to something or someone”; “I tried to make a story out of the words”).

[1] The age range and mean age of participants in the present study (65-92 years, *M* = 75.5, *SD* = 6.8) did not significantly differ from the earlier study (67-93 years, *M* = 74.5, *SD* = 6.1).

To better understand the effects of participant-chosen encoding processes, for all three encoding conditions, responses to the question “Did you use any strategy to help you learn the words during the study part of each session?” on the post-training questionnaire were coded as a) shallow (focusing on surface features of the words, rehearsal only, or vague descriptions such as “concentration” or “used good judgment”), b) intermediate (using mental imagery), or c) deep (incorporating the words into a sentence or story, self-referential processing, creating associations to other words, or using other semantically-based strategies). Coding was done by two independent raters and any discrepancies were resolved by discussion and arriving at consensus. If a participant reported multiple strategies, the one with the “deepest” level of encoding was counted.

Collapsing across condition assignment, deep encoding strategies were reported most frequently (n = 65), followed by shallow encoding strategies (n = 17). Strategies that fell at an intermediate level of encoding were rare (n = 8), and the performance of the participants who reported them tended to lie between those in the other two groups. For clarity of interpretation, our analyses below exclude these few participants and focus only on the comparison of participants who reported deep or shallow encoding[2].

The maximum lag level achieved by the last day of training did not significantly differ between participants who reported using deep (M = 30.1) versus shallow (M = 23.3) encoding ($t(80) = 1.58$, $p = .12$, $d = .43$). However, a marginally significant Day (1,2,3,4,5,6,7) × Encoding Level (shallow, deep) interaction showed a trend for deep

[2] These groups did not differ on demographic variables including age, gender, years of education, or ERVT score (all ps > .21). However, participants who reported using shallow encoding had significantly higher (i.e., worse) SBT scores (M = 2.24) than participants who reported using deep encoding (M = 1.20), $t(80) = 2.24$, $p < .03$, $d = .55$.

encoders to reach the maximum possible lag earlier in the training sessions: $F(2,158) = 2.63$, $p = .08$, $\eta_p^2 = .03$. This pattern was also evident in the number of sessions that participants required to achieve their individual maximum lag levels: Deep encoders reached asymptote earlier ($M = 5^{th}$ training day) than shallow encoders ($M = 6^{th}$ training day); $t(80) = 2.24$, $p < .03$, $d = .64$. Thus, training-task progress benefitted from deep encoding strategies, whether they were explicitly instructed (Integrated Sentences), discovered under open-ended conditions with environmental support (Strategy Choice), or covertly implemented in spite of an enforced shallow encoding task (Enforced Rehearsal). Within the self-reported deep encoders, neither the maximum lag level achieved by the last day of training nor the number of sessions required to achieve individual maximum lag levels significantly differed by encoding condition assignment (both *F*s < 1).

Contributions at retrieval

Although the focus of our experimental manipulation was on encoding processes, response times (RTs) for correct rejections of lure items during the training task test phase also reveal effects on the recollective processes engaged at retrieval. Decreases in the RT difference between short-lag and long-lag repeated lures would suggest that training helped participants restrict memory access to target information, perhaps as a result of improved encoding processes. In this “source-constrained retrieval” mode (Jacoby et al., 2005), the lag interval between repetitions should have little effect on correct rejection RTs, as efficiency in rejecting long-lag items is enhanced.

A Condition (Integrated Sentences, Strategy Choice, Enforced Rehearsal) × Day (First, Last) × Item Type (new, short-lag, long-lag) ANOVA on the RTs for correctly-rejected items revealed a significant Day × Item Type interaction, $F(2, 134) = 8.02$, $p < .005$, $\eta_p^2 = .08$, that did not interact further with Condition, $p > .17$. Follow-up analyses showed that while participants were initially slower on new items than short-lag items (1901 ms vs. 1755 ms), RTs for the two item types were equivalent by the last day of training (1530 ms vs. 1525 ms). Within the repeated item type, however (and replicating our earlier analyses; Lustig & Flegal, 2008), the differential time needed to reject long- versus short-lag items significantly decreased from the first day (2088 ms vs. 1755 ms) to the last day (1612 ms vs. 1525 ms), $F(1, 87) = 33.84$, $p < .001$, $\eta_p^2 = .28$. In other words, although response time generally improved from the first to last day of training, the largest increase was in the efficiency to reject long-lag items[3].

Participant-chosen Encoding Level also did not affect the speed at which repeated versus new lures were correctly rejected, $F < 1$. However, when comparing the RTs for correct rejections of short- versus long-lag items, there was a marginal Encoding Level (shallow, deep) × Day (First, Last) × Item Type (short-lag, long-lag) interaction, $F(1,80) = 3.31$, $p = .07$, $\eta_p^2 = .04$. Increases in the efficiency of rejecting long-lag items across training sessions were somewhat more pronounced for participants who self-reported deep encoding strategies (2180 ms on Day 1 vs. 1666 ms on Day 8, as opposed to 1805 ms on Day 1 vs. 1463 ms on Day 8 for shallow encoders).

[3] Importantly, this did not reflect an increasing bias to reject items as unstudied. Speed to accept studied items also increased from the first to last day of training, (1794 vs. 1540 ms, $t(89) = 6.96$, $p < .001$, $d = .37$), and accuracy remained stable ($M = .85$ at both occasions). In addition, overall better accuracy on studied items correlated with better training rank for all groups, whether split by Condition or Encoding Level (rs > .40, ps < .03), indicating that training-task performance was related to good encoding of the studied items.

In addition to the speed of rejecting the different lure types, differences between participants might also arise in how they processed the test phase feedback. In our earlier studies, we found that those participants who spent a proportionally longer time viewing feedback screens that followed incorrect responses than those that followed correct responses ((incorrect feedback RT - correct feedback RT)/(incorrect feedback RT + correct feedback RT) had better training ranks. This effect was replicated here for all three training conditions (Figure 2, Row 3), although it was numerically weaker in the Integrated Sentences condition than in the others. When the data were analyzed by participant-chosen Encoding Level, deep encoders also showed this correlation (r = .58, p < .001), but it was reduced and not statistically significant for shallow encoders (r = .31, p = .23). This pattern is consistent with our interpretation that participants who used shallow encoding were less likely to self-initiate controlled processing, even in response to feedback. Although no significant differences were observed across experimenter-determined training condition, participants who self-reported deep encoding strategies spent proportionally more time viewing feedback after incorrect than correct responses (M = 0.35) compared to participants who self-reported shallow encoding strategies (M = 0.25), t(80) = 2.23, p < .03, d = .66.

Transfer task performance: Changes from pre- to post-test

Shopping list. When the data were analyzed by Condition, there was a significant *negative* main effect of pre- versus post-test, F(1,87) = 10.21, p < .005, η_p^2 = .11, but no interaction with Condition, F < 1. This indicates that overall recognition accuracy was lower on the second (posttest) administration of this transfer task than on the first (pretest). This drop in accuracy suggests that the interference manipulation introduced at posttest was successful: In order to assess participants' ability to reject familiar but currently-irrelevant items, half of the lure items on the second administration were studied items on the first administration (i.e., they had been part of the shopping list on Day 1), and the other half were completely new. However, participants were instructed to identify as "studied" only list items from that day (i.e., Day 8), and to reject all other items (both previously-studied and never-studied lures). A closer examination of the data supported the role of proactive interference in reducing performance: At posttest, accuracy for items studied at pretest and then used as (familiar) lures was worse than for new, never-studied lures, $t(89) = 3.42$, $p < .005$, $d = .34$, and also worse than pretest lure item accuracy ($t(87) = 4.38$, $p < .001$, $d = .52$). Accuracy on studied items did not change between the two administrations, $F < 1$, and correct rejection of new lure items decreased, but not significantly, $t(87) = 1.76$, $p = .08$, $d = .18$ (Table 3). There were no interactions with Condition on any of these measures.

When the data were analyzed by participant-chosen Encoding Level, there was a trend for a Day (pretest, posttest) × Encoding Level (shallow, deep) interaction for overall accuracy, $F(1, 79) = 3.55$, $p = .06$, $\eta_p^2 = .04$. This interaction was significant for the correct recognition of studied items, $F(1, 79) = 4.83$, $p = .03$, $\eta_p^2 = .06$, and the correct rejection of new lure items, $F(1, 79) = 4.66$, $p = .03$, $\eta_p^2 = .06$. In both cases, participants who reported using shallow encoding showed greater declines than those who reported using deep encoding; the latter group showed little or no change (Table 4). Interestingly, the difference between correct rejection of completely new versus previously-studied lure items trended in the opposite direction, $F(1, 80) = 3.04$, $p = .09$, $\eta_p^2 = .04$, such that shallow-encoders had generally poor performance that did not differ between the item types, whereas deep-encoders performed worse on previously-studied than on never-studied lures, $t(64) = 4.21$, $p < .001$, $d = .48$.

One possible explanation for these unexpected declines in performance is that training led participants to use ineffective encoding processes, an obviously undesirable result. Another, more optimistic interpretation is that the declines in performance reflected proactive interference from the pretraining list to the posttraining list, and that the different encoding methods participants adopted during training and during the shopping-list transfer task influenced their vulnerability to this interference. That is, deep encoding may have helped preserve performance for studied and new items, with the downside of greater difficulty in rejecting familiar but currently-irrelevant lures.

Face-name recall. Face-name recall accuracy showed a significant improvement overall, $F(1,87) = 12.33$, $p < .001$, $\eta_p^2 = .12$, but did not interact with Condition, $F < 1$ (Table 3). When the data were analyzed by participant-chosen Encoding Level, deep encoders showed significant improvements from pre- to post-test, $t(64) = 3.57$, $p < .005$, $d = .45$, whereas shallow encoders did not, $t < 1$ (Table 4). However, the interaction was not statistically significant, $F(1,80) = 2.49$, $p = .12$, $\eta_p^2 = .03$. Correlations between face-name recall improvement and training rank were not significant for either group, both *r*s $< .15$.

Nonverbal, nonmemory tests (negative controls). The nonverbal, nonmemory tests used as negative controls for practice effects (Pattern Comparison, Word SOPT, Pattern SOPT) did not show any differences from pre- to post-test or any interactions with Condition, all *F*s < 1 (Table 3), consistent with the prediction that little or no improvement would be found on transfer tasks that do not emphasize the semantically-based or integrative processes targeted by our training procedure. A general increase in speed from pre- to post-test was found on the Trail-Making Test, for both Part A, $F(1, 86) = 6.95$, $p < .02$, $\eta_p^2 = .08$, and Part B, $F(1, 86) = 12.50$, $p < .005$, $\eta_p^2 = .13$, but neither interacted with Condition, both *p*s > .30. The same patterns were found when these tests were analyzed by Encoding Level (Table 4).

Notably, the Trail-Making Test was the only nonverbal test for which we did not have alternate forms, suggesting that the improvements seen here were simple practice effects. This may also explain why Jennings et al. (2005) found improvements on SOPT after repetition-lag training where we did not: Their experiment used the same forms on both pre- and post-test, whereas we used alternate forms that reduced the influence of general practice or familiarity. (See also Stamenova et al., 2014, for evidence suggesting that retrieval-focused training may be less effective in promoting transfer).

Posttest-only tasks

Surprise lure test. After the final study-test cycle on the last day of training, participants were administered a surprise old/new recognition test for the items seen in that cycle. On this test, they were asked to respond “old” to any word that had appeared in the previous training cycle, regardless of whether it had been one of the words on the studied list or one of the unstudied lures from the test phase. This test was intended to be diagnostic of how encoding strategies (experimenter-directed or participant-chosen) might influence retrieval strategies during the training task. In particular, if the Integrated Sentences condition was successful in leading participants to create a strong, integrated representation of the studied list, and thus quickly reject lure items during the test phase, then they should show relatively poor memory for those items as compared to studied items on this surprise test.

Unsurprisingly, when the data were analyzed by Condition, participants in all three groups recognized proportionally more studied items than previous-lure items from the final training session ((studied - lure) / (studied + lure)). As predicted, this difference was greater in the Integrated Sentences condition than in the other two conditions, $F(2, 78) = 3.90$, $p < .03$, $\eta_p^2 = .09$. The groups did not differ in their correct rejection of new items, or in accuracy for studied items, *F*s < 1 (Table 5). No differences were found when performance was analyzed by participant-chosen Encoding Level, all *p*s > .20.

Word and source memory. Because the different training conditions emphasized different aspects of word encoding, differences between them should be expected for word memory but not for source memory (which was not targeted for training in any of the conditions). This was indeed the case: When the data were analyzed by Condition, the three groups significantly differed on item recognition (word memory; $F(2, 79) = 3.92$, $p < .03$, $\eta_p^2 = .09$), with no group differences in source memory, $F < 1$. Post hoc *t* tests showed that word memory accuracy in the Integrated Sentences condition ($M = 72\%$) was lower than in the Strategy Choice ($M = 77\%$; $p = .08$) or Enforced Rehearsal ($M = 79\%$; $p < .01$) conditions (Table 5). This difference may have arisen if self-generated strategies (from the Strategy Choice and Enforced Rehearsal conditions) were more likely to be transferred to the word memory task than experimenter-provided strategies (from the Integrated Sentences condition), especially since the per-word presentation times in the word memory transfer task were much shorter than in the training task.

Consistent with this explanation, when the data were analyzed by participant-chosen Encoding Level, there was a marginal effect favoring deep encoders for word memory, $F(1,74) = 3.23$, $p = .08$, $\eta_p^2 = .04$. When participants in the Integrated Sentences condition were eliminated from this analysis (because the short presentation times likely discouraged use of the time-consuming sentences strategy), the effect grew stronger, $F(1,47) = 4.78$, $p < .03$, $\eta_p^2 = .09$. Encoding Level did not affect source memory accuracy regardless of whether or not the Integrated Sentences condition was included, all *p*s > .21.

Correlations between training rank and word memory accuracy were high for all three experimenter-determined encoding conditions (*r*s > .40, *p*s < .04). There were also strong correlations between training rank and source memory accuracy for participants in the Integrated Sentences ($r = .55$, $p < .01$) and Enforced Rehearsal ($r = .64$, $p < .001$) conditions, with a weaker relationship for Strategy Choice participants ($r = .29$, $p = .14$). For the Strategy Choice condition, a significant correlation between rank and word memory, with no correlation between rank and source memory, replicates the pattern observed at n = 16 in our earlier analyses (Lustig & Flegal, 2008). It may reflect a tendency for these participants to transfer the successful strategies they generated during training to a novel word memory task. In contrast, source memory, is not likely to benefit from transferred strategies because it was not a focus of the training intervention.

When the data were split by participant-chosen Encoding Level, for deep encoders training rank was a strong predictor of both word and source memory (r values of .49 and .48, respectively, $p < .001$). For shallow encoders, training rank was a strong predictor of word accuracy, $r = .57$, $p < .03$, but not source accuracy, $r = .32$, $p = .22$. Consistent with the conclusions drawn from the ANOVA results, dropping Integrated Sentences participants from the analysis slightly increased rank-word accuracy correlations for both deep and shallow encoders ($r = .54$ and $r = .67$) with little effect on source-memory correlations ($r = .41$ and $r = .29$).

<u>Self-report measures</u>

<u>Everyday Memory Questionnaire (EMQ)</u>. When the data were analyzed by Condition, there was an overall reduction in self-reported everyday memory errors from pre- to post-test, $F(1, 87) = 39.81$, $p < .001$, $\eta_p^2 = .31$, but the size of the reduction was smaller for the Integrated Sentences condition than for the other two groups, $F(2,87) = 3.24$, $p < .05$, $\eta_p^2 = .07$ (Table 3). Participant-chosen Encoding Level did not interact significantly with the reduction in everyday memory errors, $F(1, 80) = 1.72$, $p = .19$. Unlike the results for word memory (see above), eliminating the Integrated Sentences participants from the analysis did not change this pattern. These results suggest that allowing participants to choose and practice their own strategies during the training period (even covertly, if task demands imposed by the experimenter-prescribed strategy are relatively low, as in our Enforced Rehearsal condition) was more effective in reducing real-world, everyday memory errors than was prescribing a specific strategy. However, in everyday life (where participants presumably have more control over their environment and learning conditions than in the lab), the exact nature of the self-

selected strategy (i.e., involving a deep or shallow level of encoding) may not have had much influence. In other words, practicing a strategy that the participant was comfortable with and preferred to use appeared to be the important factor for improving everyday memory.

The number of EMQ errors on Day 1 (pretraining) was significantly correlated with training rank only for participants in the Strategy Choice condition (Figure 3, Row 1), replicating our previous findings (Lustig & Flegal, 2008). Also replicating our previous results, training in the Strategy Choice condition eliminated the rank-EMQ correlation on Day 8 (Figure 3, Row 2), suggesting that those participants who had the most EMQ errors on Day 1 showed the largest benefits from training. The opposite pattern of results was found for the Integrated Sentences condition, where rank did not correlate with EMQ on Day 1, but did on Day 8 (Figure 3, Rows 1 & 2). This suggests that in the Integrated Sentences condition, participants who showed the smallest training gains in the laboratory also showed little benefit in everyday memory by the end of training, perhaps because the experimenter-prescribed strategy was a poor match for their abilities and preferences. Of note, a significant correlation ($r = .65$, $p < .05$) between training rank and proportional change on the EMQ for the Strategy Choice condition at $n = 16$ in our earlier analyses (Lustig & Flegal, 2008) still trends in the same direction but is no longer significant ($r = .27$, $p = .15$) with the complete sample of $n = 30$ (Figure 3, Row 3). Removing one Strategy Choice participant with an outlying EMQ-change score (i.e., p(EMQ change) = -.44 for this individual, compared to $M = .46$, $SD = .35$, for the entire group) restores the significance of the correlation for this condition, $r = .39$, $p < .05$.

When the data were split by participant-chosen Encoding Level, deep encoders showed a small correlation between training rank and EMQ errors on Day 8 (r = .29, p < .05, those with worse training ranks reported more errors), but no correlations with Day 1 errors or with the proportional change in errors over training (both rs < .15, ps > .30). Inspection of the data suggested that the Day 8 correlations may have been due to two outlying participants. For shallow encoders, there was no relationship between training rank and the Day 1 or Day 8 errors, but a significant correlation between training rank and EMQ change (r = .51, p < .05). This suggests that participants who failed to self-initiate successful encoding strategies over the course of training were more likely to have entered the study with a large number of reported memory errors on the first administration of the EMQ, with more room for improvement after training.

Memory Self-Efficacy Questionnaire (MSEQ). Given that the EMQ is a self-report measure, a potential concern is that improvements on it may reflect a placebo effect. That is, did participants report fewer memory errors because they knew that they were in a memory-training study and thus thought they "should" show a better memory in their everyday lives? The differences in improvement for the three encoding conditions, described above, argue against this interpretation. To further examine this possibility, we also examined changes in the MSEQ. If fewer reported everyday memory errors on the EMQ reflected a placebo effect, then confidence in memory ability should likewise show an increase, and should be correlated with improvements on the EMQ.

This was not the case. When the MSEQ data were analyzed using a Day (pretest, posttest) × Condition repeated-measures ANOVA, there was no main effect for either factor, both ps > .25. The interaction was marginally significant, F(2, 87) = 2.31, p

= .11, η_p^2 = .05. Interestingly, although Enforced Rehearsal participants were assigned to a condition designed to suppress strategies that would improve memory performance, they showed an increase in MSEQ, $t(29) = 2.46$, $p < .03$, $d = .28$, that was not shared by the other two groups, both *t*s < 1. This effect appears to be driven by those Enforced Rehearsal participants who self-reported deep encoding strategies, $t(17) = 3.13$, $p < .01$, $d = .36$, as MSEQ scores did not increase significantly for the other participants in the Enforced Rehearsal condition ($t < 1$). Speculatively, the deep encoders may have experienced a boost in memory self-efficacy from successfully engaging covert, associative encoding processes at the same time as overtly following the less effective experimenter-provided strategy, and discovering a resultant benefit to memory performance.

Importantly, EMQ scores (Day 1, Day 8, and proportional change) did not differ between the Enforced Rehearsal participants who did vs. did not report using deep encoding, arguing against a change in self-efficacy accounting for fewer reported EMQ errors. In addition, when the data were collapsed across condition assignment, there were no MSEQ differences related to self-reported encoding level, $F < 1$. Furthermore, there were no correlations between proportional change on the EMQ and on the MSEQ, regardless of whether data were analyzed over all groups, split by Condition, or split by participant-chosen Encoding Level (all *r*s < .20, *p*s > .30). Thus, there is no evidence to suggest that changes in perceived memory efficacy that might potentially be caused by a placebo effect of training contributed to the reduction in self-reported memory errors on the EMQ.

Discussion

The present results confirm many of the preliminary conclusions from our earlier analyses (Lustig & Flegal, 2008), and extend them with the addition of data from the Enforced Rehearsal encoding condition. In particular, the large size of our complete sample (n = 30 per condition) helped to reveal important differences between the effects of experimenter-directed and participant-chosen encoding strategies. The Integrated Sentences strategy produced large training gains only for participants who had strong verbal skills to begin with, and although a surprise lure test given on the last day of training showed that Integrated Sentences participants benefitted from creating integrated representations of studied items, their lower accuracy on a posttraining item recognition test suggested that benefits do not transfer when task demands discourage a time-consuming encoding strategy. In contrast, the Strategy Choice condition minimized the influence of pre-existing ability on training gains, showed a relationship between training rank and self-reported everyday memory errors, and facilitated participant-driven deep encoding strategies which were associated with positive training and transfer effects.

Relative to a previous experiment in which encoding was self-paced and unconstrained (Bissig & Lustig, 2007), enforced study times in the present study benefitted training-task performance in all three experimenter-determined encoding conditions. Even in the Enforced Rehearsal condition, which was originally designed to serve as a control condition that would suppress deep encoding processes, sufficient study time was evidently available for many participants to covertly engage in deep processing while also completing the overt rehearsal task (as discussed below). As a result, environmental support to facilitate the self-initiation of effortful memory processes diminished the association between older age and poorer training task performance that was present under open-ended conditions (Bissig & Lustig, 2007). However, training-task success was significantly correlated with vocabulary and years of education only for participants in the Integrated Sentences condition, suggesting that the verbal encoding strategy was most beneficial for participants who already possessed strong verbal skills.

Encoding condition assignment also influenced performance on self-report measures and unpracticed transfer tasks. In the Strategy Choice condition, training-task rank was correlated with the number of memory errors reported on the EMQ on Day 1, but not on Day 8 – suggesting that participants in this condition who experienced the most memory errors pre-training experienced the largest benefits from training. In contrast, in the Integrated Sentences condition, training-task rank was correlated with the number of EMQ memory errors on Day 8, but not on Day 1 – suggesting that participants in this condition who showed the smallest training gains in the laboratory also experienced little benefit in the reduction of memory errors post-training, perhaps because the compulsory verbal strategy was a poor match for their abilities and preferences.

In the surprise Lure Test that followed the final training session, the proportion of recognized studied items to previous-lure items was greatest in the Integrated Sentences condition, suggesting that an experimenter-enforced strategy to increase deep, associative processing at encoding produces a benefit of decreased lure processing at retrieval. However, a cost of the Integrated Sentences strategy was evident in poorer item recognition on the posttraining word and source memory task, relative to the other two conditions, indicating that training in such a time-dependent manner does not benefit (and may even impair) performance in a situation where exposure to the stimuli was time-limited.

Taken together, these results are consistent with the hypothesis that more benefits and fewer costs accrue when environmental support is provided but participants are allowed to generate their own encoding strategies. Although the processing required by the experimenter-provided Integrated Sentences instructions might be characterized as deep or elaborative, this strategy apparently led to different results than did participant-chosen deep encoding methods. In the absence of explicit instructions to perform an integrative encoding task, older adults may be more likely to generate strategies which emphasize item-specific processing more than cross-item connections, a tendency that is consistent with associative-deficit theories of cognitive aging (Naveh-Benjamin, 2000) as well as previous empirical findings (e.g., Luszcz, Roberts, & Mattiske, 1990). Self-reported level of encoding, independent of experimenter-prescribed encoding condition, was associated with training-task success. Participants whose post-training questionnaire responses indicated that they used deep encoding strategies advanced through the levels of the repetition-lag task more quickly than participants who reported using shallow encoding strategies. On the posttraining word and source memory task, a differential strategy benefit was also found in that training-task rank predicted source accuracy for deep encoders, but not shallow encoders.

Although their central question, methods, and outcome measures were quite different from ours, the results of Kirchhoff and colleagues (2012a; 2012b) also support the idea that conditions encouraging older adults to engage in effective encoding but allowing them to choose their own methods of doing so can be very effective. Prior to training, older adults were much more likely than young adults to report using “no strategy” at encoding. After exposure to three different semantic encoding strategies (pleasantness ratings, self-relevance, and sentence generation), older adults decreased their “no strategy” reports and increased their use of semantic encoding, so that overall their rate of using deep encoding strategies was at least as high as that of young adults (Kirchhoff et al., 2012a).

Notably, in their methods Kirchhoff et al. (2012a) assumed the hypothesis tested here, that older adults “would be most likely to self-initiate self-selected semantic encoding strategy(ies)” (pg 791). The approximately equal increase in self-reported use across their different strategy types suggests that some strategies were indeed a better fit for some individuals than others. An earlier study with young adults provides additional support for that view, as which strategy participants used was less predictive of later memory than the degree to which they activated brain regions associated with that strategy (Kirchhoff & Buckner, 2006). The strategy training study conducted by Kirchhoff et al. (2012a; 2012b) showed that training increased older adults’ activation of prefrontal regions associated with self-initiated processing as well as their subsequent memory, and activation increases correlated with increased performance.

Finally, a later analysis of the retrieval data found that the benefits of increased semantic encoding selectively increased recollection and decreased reliance on familiarity, and increased activation in hippocampal regions associated with recollection (Kirchhoff et al., 2012b). Likewise, both our present results and previous results (Bissig & Lustig, 2007; Lustig & Flegal, 2008), indicate that improving encoding helped older adults to reject familiar but unstudied lures in the repetition-lag procedure. In short, although Kirchhoff and colleagues assumed rather than directly tested the importance of participant-chosen strategies, and used neural rather than behavioral transfer measures, their results converge with ours to suggest that providing older adults with both support and freedom for deep encoding encourages their later self-initiated use of such strategies, and that this has beneficial effects for recollection (cf., Bailey, Dagenbach, & Jennings, 2011).

The findings of Kirchhoff et al. (2012a; 2012b) are somewhat ambiguous as to whether older adults' initial failures to use successful encoding strategies reflected a lack of knowledge about those strategies versus a failure to engage them. Our findings, together with the literature on age deficits in metacognitive monitoring, point to the latter explanation. The Strategy Choice condition in the present study differed from the Open-Ended condition in a previous experiment (Bissig & Lustig, 2007) only in the environmental support provided by fixed encoding time, but this factor proved to be sufficient to diminish age and ability differences in training task performance that were present when encoding time was self-paced. Deficient monitoring skills may explain why older adults can benefit from extended time, as our results and others have shown (e.g., Murphy, et al., 1981; Paxton, et al., 2006), yet are unlikely to allocate as much time as they need when left to their own devices (Bissig & Lustig, 2007; Dunlosky & Connor, 1997; Murphy, et al., 1981). This interpretation was supported by a study in which monitoring training combined with memory strategy training produced associative memory improvements for older adults, relative to strategy training alone – when the study phase was self-paced, but not experimenter-paced (Dunlosky et al., 2003). Similarly, an early study (Murphy et al., 1981) showed that older adults given enforced encoding time but no strategy training actually performed better on a serial recall task than older adults provided with explicit strategy training, suggesting that an age-related monitoring deficit impedes efficient (and sufficient) use of study time under self-paced conditions.

Hertzog, Price, and Dunlosky (2012) gave older and young adults practice that was either supervised (i.e., experimenter-mandated, using rote repetition and interactive imagery strategies) or unsupervised (i.e., freely-chosen, following exposure to a list of potential strategies) in an associative memory task with fixed encoding time. For a second study period, all participants were free to choose their own encoding strategies. Young adults' recall performance improved (reflecting knowledge updating and a shift toward using more effective strategies) while older adults' did not (indicating a tendency to stick with previous, suboptimal strategies). This result might be considered another example of deficient monitoring skills, with older adults failing to monitor the efficacy of different memory strategies and regulate their behavior accordingly (see also Brigham & Pressley, 1988).

It is worth noting that the memory strategy use reported in Hertzog, Price, and Dunlosky (2012) was freely-chosen, but not strictly spontaneous. The authors acknowledge that pre-exposure to a list of potential strategies in their experiment may have prompted some participants in the unsupervised learning condition to start off with more successful strategies (e.g., interactive imagery) than they would have otherwise thought to use on their own (cf. Dunlosky & Hertzog, 2001). An open question, then, is what determines how and when older adults will spontaneously produce effective memory strategies? Evidence suggests that self-generated strategies are at least as beneficial for older adults as mnemonics provided by an experimenter (Baltes, et al., 1989; Derwinger, et al., 2003; Hill, et al., 1990). Derwinger and colleagues (2005) trained older adults in a number memory task and followed up after a delay of 8 months. Differences in posttraining memory performance emerged when environmental support was withdrawn: Accuracy declined for a group that learned a mnemonic strategy, but improved for a group that had been allowed to choose their own strategies. A possible explanation for these maintenance effects is that older adults who develop their own strategies may be more likely to use them in everyday life – in effect, continuing to "train" even after the formal intervention has ended.

Furthermore, self-generated strategies are presumably more likely to be personally tailored to an individual's strengths than a one-size-fits-all approach. Evidence from the present study supports this interpretation, showing that pre-existing verbal ability predicted training gains in the Integrated Sentences condition, but not in the Strategy Choice condition (see Figure 2, Row 2). This suggests that older adults with lower verbal skills were not able to derive as much benefit from an experimenter-provided encoding strategy that emphasized verbal processing. With environmental support to encourage sufficient attention and effort at encoding, however, these same older adults may have been able to make use of abilities from other domains they would not have otherwise brought to bear. In post-training questioning, successful participants from the Strategy Choice condition typically reported using deep encoding strategies, although not always involving verbal processing. Therefore, enforced study time alone appears to have been effective in directing older adults to self-initiate controlled cognitive processes.

However, the influence of pre-existing ability cuts both ways. Individuals with lower initial ability levels may be most in need of intervention, yet typically show the least improvement in conventional training programs (Verhaeghen & Marcoen, 1996; Verhaeghen, et al., 1992; Yesavage, et al., 1990). In the present study, even when environmental support in the form of long encoding times was available, there was variability in the efficacy of participant-chosen encoding strategies. In post-training questioning, a subgroup of participants in the Strategy Choice condition reported using suboptimal encoding methods involving relatively shallow levels of processing, rote rehearsal, or non-specific effort-based strategies – and these participants did not benefit from training to the same extent as their peers who engaged more effective encoding methods under the very same unsupervised learning conditions. An important question for future memory training research will be how to design interventions that promote the discovery and use of successful memory strategies but allow enough flexibility for training plans to be customized to a wide range of abilities and preferences.

In summary, the practical implications of our results and previous studies are that cognitive training for older adults may work best when it adopts a “personalized medicine” approach, balancing structured environmental support that cues the participant to engage encoding processes with sufficient open-endedness to allow the participant to choose the processes that are most comfortable and work best for them.

Especially when working with older adults that range in ability, introducing different strategies can be helpful, but care must be taken especially for low-ability older adults that the strategy itself is not too challenging for them to master. Our results from the Strategy Choice condition suggest that such a balance may also improve the transfer of training to other laboratory tasks and to everyday life, though we suspect that on this front there is room for improvement in future training studies. In particular, it may be useful to use training materials (not just transfer tasks) with more ecological validity or that at least remind older adults of real-world situations (e.g., grocery lists, medications), so that they may serve as cues to engage encoding when similar materials are encountered in everyday life (see Jobe et al., 2001, for a step in this direction). Successful cognitive training and transfer of trained abilities to real-life improvement for older adults remains a challenging translational problem, but taking the balance between support and open-endedness into account may be an important step towards its resolution.

Acknowledgments

We thank the many dedicated research assistants of the Lustig lab for their help on this project. We are especially grateful to Jennifer Humfleet and Caitlin Mallory for stimulus preparation and data entry, Taryn Weil, Lauren Schiff, Monica Kumar, Monika Potrykus, and Phuoc Anh Nguyen for participant scheduling, Yosi Nacach and Elyssa Guslits for coding self-report strategy data, and Katherine Cost and Mary Winters for superb lab management. We also thank the Healthier Black Elders Center at Wayne State University for their assistance with participant recruitment. Portions of this work were presented at the Annual Meeting of the Cognitive Neuroscience Society, April 2010, in Montreal.

This work was funded by the National Institute on Aging Grants R03 AG029329 awarded to Cindy Lustig and P30 AG024824 awarded to the University of Michigan Older Americans Independence Center.

References

Armitage, S. G. (1946). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychological Monographs, 60*(1), 1-48.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*(6), 452-471.

Bailey, H., Dagenbach, D., & Jennings, J. M. (2011). The locus of the benefits of repetition-lag memory training. *Aging, Neuropsychology, and Cognition, 18*(5), 577-593.

Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., Morris, J. N., Rebok, G. W., Smith, D. M., Tennstedt, S. L., Unverzagt, F. W., & Willis, S. L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *JAMA: Journal Of The American Medical Association, 288*(18), 2271-2281.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*, 445-459.

Baltes, P. B., Sowarka, D., & Kliegl, R. (1989). Cognitive training research on fluid intelligence in old-age: What can older adults achieve by themselves? *Psychology and Aging, 4*(2), 217-221.

Berry, J. M., West, R. L., & Dennehey, D. M. (1989). Reliability and validity of the Memory Self-Efficacy Questionnaire. *Developmental Psychology, 25*(5), 701-713.

Bissig, D., & Lustig, C. (2007). Who benefits from memory training? *Psychological Science, 18*(8), 720-726. doi: doi:10.1111/j.1467-9280.2007.01966.x

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake & J. N. Towse (Eds.), *Variation in working memory* (pp. 76–106): Oxford University Press.

Brigham, M. C., & Pressley, M. (1988). Cognitive monitoring and strategy choice in younger and older adults. *Psychology and Aging, 3*(3), 249-257.

Calero, M. D., & Navarro, E. (2007). Cognitive plasticity as a modulating variable on the effects of memory training in elderly persons. *Archives of Clinical Neuropsychology, 22*(1), 63-72. doi: 10.1016/j.acn.2006.06.020

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 191-211). New York: Plenum Press.

Craik, F.I.M. & Rose, N.S. (2012). Memory encoding and aging: A neurocognitive perspective. *Neuroscience & Biobehavioral Reviews, 36*(7), 1729-1739.

Craik, F. I. M., & Rabinowitz, J. C. (1985). The effects of presentation rate and encoding task on age-related memory deficits. *Journal of Gerontology, 40*(3), 309-315.

Derwinger, A., Stigsdotter Neely, A., & Bäckman, L. (2005). Design your own memory strategies! Self-generated strategy training versus mnemonic training in old age: An 8-month follow-up. *Neuropsychological Rehabilitation, 15*(1), 37-54.

Derwinger, A., Stigsdotter Neely, A., Persson, M., Hill, R. D., & Bäckman, L. (2003). Remembering numbers in old age: Mnemonic training versus self-generated strategy training. *Aging, Neuropsychology, and Cognition, 10*(3), 202-214.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory and Cognition, 25*(5), 691-700.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: What is the role of strategy production? *Psychology And Aging, 13*(4), 597-607.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory and Cognition, 29*(2), 247-253.

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology, 41*(2), 389-400. doi: 10.1037/0012-1649.41.2.389

Dunlosky, J., Kubat-Silman, A. K., & Hertzog, C. (2003). Training monitoring skills improves older adults' self-paced associative learning. *Psychology and Aging, 18*(2), 340-345. doi: 10.1037/0882-7974.18.2.340

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician *Journal Of Psychiatric Research, 12*(3), 189-198.

Froger, C., Bouazzaoui, B., Isingrini, M., & Taconnat, L. (2012). Study time allocation deficit of older adults: The role of environmental support at encoding? *Psychology and Aging, 27*(3), 577-588.

Hertzog, C., Price, J., & Dunlosky, J. (2012). Age differences in the effects of experimenter-instructed versus self-generated strategy use. *Experimental Aging Research, 38*(1), 42-62. doi: 10.1080/0361073x.2012.637005

Hill, R. D., Allen, C., & Gregory, K. (1990). Self-generated mnemonics for enhancing free-recall performance in older learners. *Experimental Aging Research, 16*(3), 141-145.

Hill, R. D., Yesavage, J. A., Sheikh, J., & Friedman, L. (1989). Mental status as a predictor of response to memory training in older adults. *Educational Gerontology, 15*(6), 633-639.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review, 12*(5), 852-857.

Jennings, J. M., & Jacoby, L. L. (2003). Improving memory in older adults: Training recollection. *Neuropsychological Rehabilitation, 13*(4), 417-440.

Jennings, J. M., Webster, L. M., Kleykamp, B. A., & Dagenbach, D. (2005). Recollection training and transfer effects in older adults: Successful use of a repetition-lag procedure. *Aging, Neuropsychology, and Cognition, 12*(3), 278-298.

Jobe, J. B., Smith, D. M., Ball, K., Tennstedt, S. L., Marsiske, M., Willis, S. L., et al. (2001). ACTIVE: A cognitive intervention trial to promote independence in older adults. *Controlled Clinical Trials, 22*, 453-479.

Jones, S., Nyberg, L., Sandblom, J., Stigsdotter Neely, A., Ingvar, M., Petersson, K. M., & Bäckman, L. (2006). Cognitive and neural plasticity in aging: General and task-specific limitations. *Neuroscience & Biobehavioral Reviews, 30*(6), 864-871.

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short orientation-memory-concentration test of cognitive impairment. *American Journal Of Psychiatry, 140*(6), 734-739.

Kausler, D. H. (1994). *Learning and memory in normal aging*. New York: Academic Press.

Kirchhoff, B. A., Anderson, B. A., Barch, D. M., & Jacoby, L. L. (2012a). Cognitive and neural effects of semantic encoding strategy training in older adults. *Cerebral Cortex, 22*(4), 788-799.

Kirchhoff, B. A., Anderson, B. A., Smith, S. E., Barch, D. M., & Jacoby, L. L. (2012b). Cognitive training-related changes in hippocampal activity associated with recollection in older adults. *NeuroImage, 62*(3), 1956-1964. doi:10.1016/j.neuroimage.2012.06.017

Kirchhoff, B. A., & Buckner, R. L. (2006). Functional-anatomic correlates of individual differences in memory. *Neuron, 51*(2) , 263-274.

Logan, J. M., Sanders, A. L., Snyder, A. Z., Morris, J. C., & Buckner, R. L. (2002). Under-recruitment and nonselective recruitment: Dissociable neural mechanisms associated with aging. *Neuron, 33*(5), 827-840.

Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training-induced compensation versus magnification of individual differences in memory performance. *Frontiers in Human Neuroscience, 6*, 141. doi:10.3389/fnhum.2012.00141

Lustig, C., & Flegal, K. E. (2008). Targeting latent function: Encouraging effective encoding for successful memory training and transfer. *Psychology and Aging, 23*(4), 754-764.

Lustig, C., & Lin Z. (in press). Memory: Behavioral and neural bases. In K.L. Schaie and S.L. Willis (Eds). *Handbook of the Psychology of Aging*.

Lustig, C., Shah, P., Seidler, R., & Reuter-Lorenz, P. A. (2009). Aging, training, and the brain: A review and future directions. *Neuropsychology Review, 19*(4), 504-522.

Luszcz, M. A., Roberts, T. H., & Mattiske, J. (1990). Use of relational and item-specific information in remembering by older and younger adults. *Psychology and Aging, 5*(2), 242-249.

Murphy, M. D., Sanders, R. E., Gabriesheski, A. S., & Schmitt, F. A. (1981). Metamemory in the aged. *Journals of Gerontology, 36*(2), 185-193.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*(5), 1170-1187. doi: 10.1037//0278-7393.26.2.1170

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging, 22*(1), 202-208. doi: 10.1037/0882-7974.22.1.202

Ossher, L., Flegal, K. E., & Lustig, C. (2013). Everyday memory errors in older adults. *Aging, Neuropsychology, and Cognition, 20*(2), 220-242.

Park, D. C., Gutchess, A. H., Meade, M. L., & Stine-Morrow, E. A. L. (2007). Improving cognitive function in older adults: Nontraditional approaches. *Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 62*, 45-52.

Paxton, J. L., Barch, D. M., Storandt, M., & Braver, T. S. (2006). Effects of environmental support and strategy training on older adults' use of context. *Psychology and Aging, 21*(3), 499-509.

Rebok, G. W., & Balcerak, L. J. (1989). Memory self-efficacy and performance differences in young and old adults: The effect of mnemonic training. *Developmental Psychology, 25*(5), 714-721.

Rebok, G. W., Carlson, M. C., & Langbaurn, J. B. S. (2007). Training and maintaining memory abilities in healthy older adults: Traditional and novel approaches.

Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 62, 53-61.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L.-G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Data from a population based study. *Psychology and Aging, 20*(1), 3-18.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*(3), 403-428.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age-differences in working memory. *Developmental Psychology, 27*(5), 763-776.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10*(4), 527-539.

Stamenova, V., Jennings, J. M., Cook, S. P., Walker, L. A. S., Smith, A. M., & Davidson, P. S. R. (2014). Training recollection in healthy older adults: Clear improvements on the training task, but little evidence of transfer. *Frontiers in Human Neuroscience, 8*, 898. doi:10.3389/fnhum.2014.00898

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 341-357.

Thompson, L. A., & Kliegl, R. (1991). Adult age effects of plausibility on memory: The role of time constraints during encoding. *Journal of Experimental Psychology: Learning Memory and Cognition, 17*(3), 542-555.

Touron, D. R., & Hertzog, C. (2004). Distinguishing age differences in knowledge, strategy use, and confidence during strategic skill acquisition. *Psychology and Aging, 19*(3), 452-466. doi: 10.1037/0882-7974.19.3.452

Verhaeghen, P., & Marcoen, A. (1994). Production deficiency hypothesis revisited: Adult age-differences in strategy use as a function of processing resources. *Aging and Cognition, 1*(4), 323-338.

Verhaeghen, P., & Marcoen, A. (1996). On the mechanisms of plasticity in young and older adults after instruction in the method of loci: Evidence for an amplification model. *Psychology and Aging, 11*(1), 164-178.

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging, 7*(2), 242-251.

Wood, L. E., & Pratt, J. D. (1987). Pegword mnemonic as an aid to memory in the elderly: A comparison of 4 age-groups. *Educational Gerontology, 13*(4), 325-339.

Yesavage, J. A., & Rose, T. L. (1984). The effects of a face-name mnemonic in young, middle-aged, and elderly adults. *Experimental Aging Research, 10*(1), 55-57.

Yesavage, J. A., Sheikh, J., Tanke, E. D., & Hill, R. (1988). Response to memory training and individual-differences in verbal intelligence and state anxiety. *American Journal of Psychiatry, 145*(5), 636-639.

Yesavage, J. A., Sheikh, J. I., Friedman, L., & Tanke, E. (1990). Learning mnemonics: Roles of aging and subtle cognitive impairment. *Psychology and Aging, 5*(1), 133-137.

Table 1
Participant demographics by condition.

	Integrated Sentences		**Strategy Choice**		**Enforced Rehearsal**	
	M	*SD*	M	*SD*	M	*SD*
Age (years)	75.7	*6.7*	75.6	*6.7*	75.2	*7.3*
Edu. (years)	16.3	*3.7*	16.7	*3.3*	16.4	*2.4*
ERVT	27.4	*11.3*	28.6	*10.7*	26.7	*11.7*
MMSE	28.4	*1.6*	28.7	*1.5*	28.0	*1.6*
SBT	1.0	*1.5*	1.9	*2.3*	1.6	*1.6*

Note: ERVT = Extended Range Vocabulary Test (maximum score = 48); MMSE = Mini Mental State Evaluation (maximum score = 30; higher scores = better performance); SBT = Short Blessed Test (maximum score = 28; higher scores = worse performance).

Table 2
List of transfer tasks.

Task	Domain
Verbal/associative memory tests	
Shopping list	Semantically-based and integrative processes
Face-name recall	targeted by encoding training
Nonverbal, nonmemory tests (negative controls)	
Pattern Comparison Test	Cognitive speed
Trail-Making Test, Part A	Cognitive speed
Trail-Making Test, Part B	Executive function
Self-ordered pointing test	Working memory
Self-report measures	
MSEQ	Memory self-efficacy
EMQ	Everyday memory errors
Posttest-only tasks	
Surprise lure test	Previous-lure item recognition
Word memory	Verbal item recognition
Source memory	Voice source recognition

Note: MSEQ = Memory Self-Efficacy Questionnaire; EMQ = Everyday Memory Questionnaire.

Table 3
Transfer task performance by condition, with p-values reported for Day x Condition interactions.

	Integrated Sentences				**Strategy Choice**				**Enforced Rehearsal**				
	Pretraining		Posttraining		Pretraining		Posttraining		Pretraining		Posttraining		
	M	*SD*	M	*SD*	M	*SD*	M	*SD*	M	*SD*	M	*SD*	p
Verbal/associative memory tests													
Shopping-list recognition accuracy (%)	95.1	*5.7*	92.8	*7.2*	96.8	*3.9*	94.8	*8.4*	96.4	*6.9*	93.4	*9.1*	.86
Studied item accuracy (%)	93.0	*12.4*	93.6	*6.3*	95.3	*6.8*	96.0	*5.7*	94.6	*6.7*	92.1	*11.0*	.46
Never-studied lure accuracy (%)	96.3	*6.9*	93.9	*9.5*	97.4	*4.8*	96.6	*6.6*	97.1	*8.2*	95.3	*10.5*	.84
Previously-studied lure (from Pretraining) accuracy (%)			90.6	*10.8*			91.5	*16.9*			92.4	*11.0*	—
Face-name association recall accuracy (%)	14.7	*19.3*	22.3	*20.3*	28.3	*24.8*	37.0	*29.8*	19.0	*16.3*	28.3	*25.1*	.96
Face-name association total accuracy (recall plus recognition; %)	89.3	*12.0*	87.7	*15.0*	88.3	*13.9*	87.3	*12.8*	91.7	*11.5*	87.7	*11.7*	.76
Nonverbal, nonmemory tests (negative controls)													
Pattern Comparison Test (no. of correct responses in 20 s)	10.1	*2.3*	9.5	*2.7*	9.7	*2.1*	10.4	*2.9*	10.0	*1.5*	9.9	*2.5*	.08
Trail-Making Test, Part A (in seconds)	37.2	*17.5*	33.9	*10.7*	40.5	*19.7*	35.8	*15.7*	33.7	*10.6*	31.9	*9.8*	.62
Trail-Making Test, Part B (in seconds)	109.9	*86.5*	83.3	*45.6*	97.2	*46.5*	83.8	*34.7*	99.7	*55.4*	90.6	*43.1*	.32
SOPT pattern (no. of unique responses)	11.6	*1.7*	11.5	*1.6*	11.9	*1.4*	12.0	*1.4*	11.7	*1.4*	12.2	*1.7*	.38
SOPT word (no. of unique responses)	14.3	*1.4*	14.3	*1.5*	14.1	*1.4*	14.5	*1.4*	13.9	*1.5*	14.1	*1.5*	.77
Self-report measures													
MSEQ overall memory self-efficacy strength	64.7	*18.5*	63.2	*15.9*	65.5	*19.4*	66.0	*17.7*	59.9	*19.3*	65.4	*20.3*	.11
EMQ number of memory errors	10.2	*8.4*	7.2	*9.6*	10.7	*7.6*	3.7	*3.8*	14.4	*13.5*	5.1	*4.5*	.04

Note: SOPT = Self-ordered pointing test (maximum score = 16); MSEQ = Memory Self-Efficacy Questionnaire (maximum score = 100); EMQ = Everyday Memory Questionnaire.

Table 4
Transfer task performance by self-reported level of encoding, with p-values reported for Day x Encoding Level interactions.

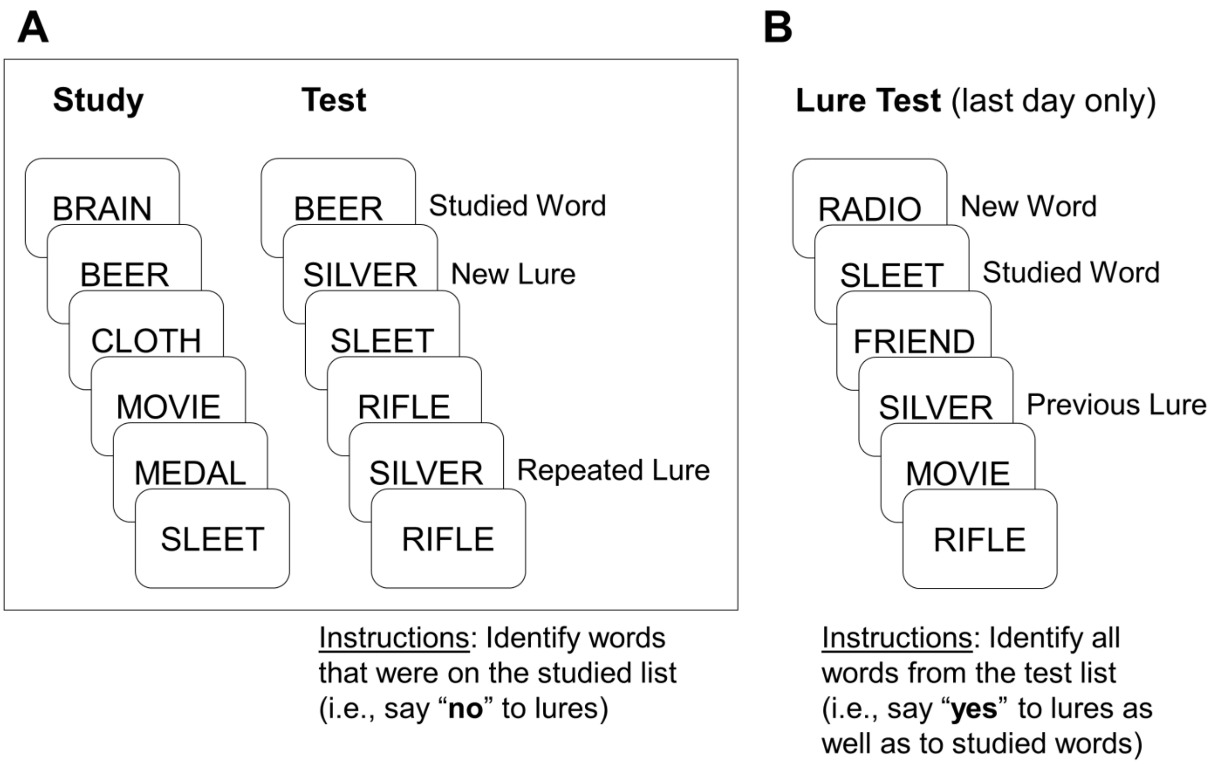
	Deep Encoding				**Shallow Encoding**				
	Pretraining		Posttraining		Pretraining		Posttraining		
	M	*SD*	M	*SD*	M	*SD*	M	*SD*	p
Verbal/associative memory tests									
Shopping-list recognition accuracy (%)	96.0	*5.8*	94.4	*7.6*	95.6	*5.8*	90.2	*9.8*	.06
Studied item accuracy (%)	94.3	*8.5*	95.4	*5.8*	94.0	*11.3*	88.2	*12.2*	.03
Never-studied lure accuracy (%)	96.8	*6.9*	96.5	*7.3*	96.2	*7.1*	91.4	*12.6*	.03
Previously-studied lure (from Pretraining) accuracy (%)			91.0	*14.3*			90.9	*9.7*	—
Face-name association recall accuracy (%)	20.6	*21.3*	31.2	*25.5*	17.1	*16.5*	17.6	*23.3*	.12
Face-name association total accuracy (recall plus recognition; %)	89.4	*12.5*	88.9	*12.0*	90.0	*13.2*	82.4	*15.6*	.10
Nonverbal, nonmemory tests (negative controls)									
Pattern Comparison Test (no. of correct responses in 20 s)	10.0	*2.1*	9.8	*2.9*	9.7	*1.3*	9.9	*2.3*	.51
Trail-Making Test, Part A (in seconds)	36.9	*15.5*	32.9	*10.8*	39.2	*19.8*	38.8	*15.8*	.25
Trail-Making Test, Part B (in seconds)	101.3	*71.1*	82.1	*40.2*	114.8	*42.6*	110.1	*40.3*	.25
SOPT pattern (no. of unique responses)	11.8	*1.4*	12.0	*1.5*	11.2	*1.6*	11.5	*2.0*	.88
SOPT word (no. of unique responses)	14.1	*1.3*	14.3	*1.5*	13.6	*1.9*	13.9	*1.3*	.79
Self-report measures									
MSEQ overall memory self-efficacy strength	64.8	*18.4*	66.0	*16.4*	57.7	*20.2*	59.2	*22.8*	.93
EMQ number of memory errors	11.0	*9.5*	5.0	*7.0*	14.6	*12.4*	5.3	*3.4*	.19

Note: SOPT = Self-ordered pointing test (maximum score = 16); MSEQ = Memory Self-Efficacy Questionnaire (maximum score = 100); EMQ = Everyday Memory Questionnaire. Encoding level defined by participants' post-training questionnaire responses (see text).

Table 5
Posttest-only task performance by condition, with p-values reported for main effect of Condition.

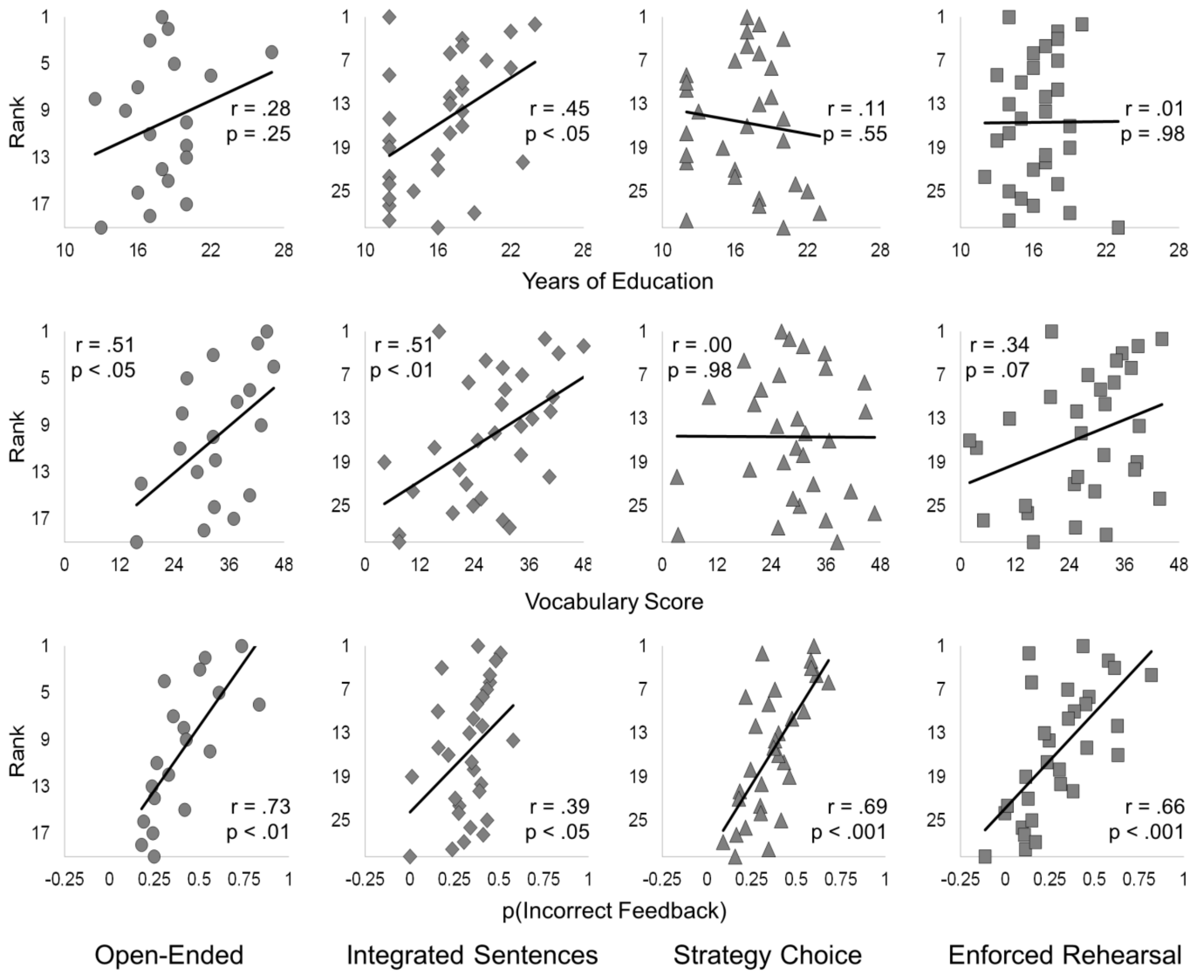
	Integrated Sentences		**Strategy Choice**		**Enforced Rehearsal**		
	M	*SD*	M	*SD*	M	*SD*	p
Lure Test accuracy							
New items (%)	88.5	*11.8*	89.9	*6.3*	89.1	*10.1*	.87
Studied items (%)	89.0	*12.7*	91.5	*13.2*	92.4	*12.3*	.63
Previous lure items (%)	50.9	*34.6*	65.3	*22.4*	63.1	*22.7*	.13
Proportion of studied items to previous-lure items	0.37	*0.08*	0.19	*0.03*	0.21	*0.04*	.02
Word memory accuracy (%)	71.8	*9.9*	76.5	*10.4*	79.3	*9.7*	.02
Source memory accuracy (%)	50.6	*17.1*	52.9	*15.8*	53.7	*16.5*	.78

Figure 1



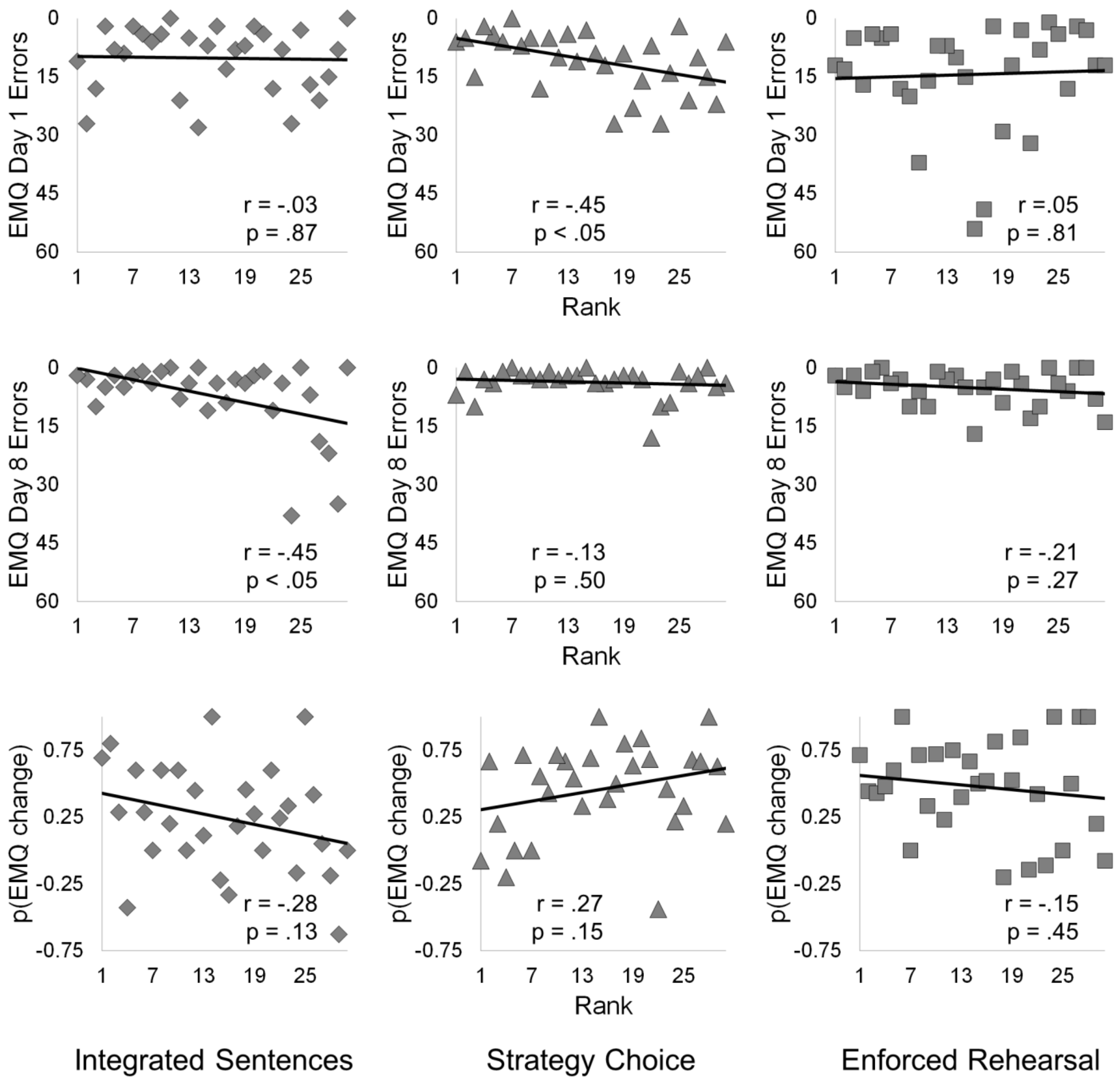
A: Training task. In the *study* phase, participants viewed 30 words presented individually for 14 seconds each; in the *test* phase, participants indicated whether each word was on the studied list or was an unstudied lure. Unstudied lures were repeated within each test list, with half of the repetitions occurring at a short lag (e.g., one intervening item between the first and second presentations of RIFLE in this example), and the other half occurring at a long lag (e.g., two intervening items between the first and second presentations of SILVER in this example). **B: Surprise lure test**. A surprise recognition test for unstudied lures from the final training test list was given on the last day of training; participants indicated whether each word was present in the final test session (regardless of whether it was originally a studied word or a previous lure item) or was a completely new word.

Figure 2



Correlations between training rank and ability measures, and attention to feedback after incorrect responses. *Note*: Open-Ended = data from Bissig & Lustig (2007), in which encoding time and strategy were unconstrained. *p*(Incorrect Feedback) = (incorrect feedback RT - correct feedback RT)/(incorrect feedback RT + correct feedback RT). RT = response time.

Figure 3



Correlations between training rank and scores from the Everyday Memory Questionnaire (EMQ). *Note*: Axes are arranged so that better scores (i.e., fewer memory errors) are always higher on the y axis. *p*(EMQ change) = (Day 1 errors - Day 8 errors)/(Day 1 errors + Day 8 errors)